An Exploratory Analysis on the Risk to be Offended on the Internet

Susanne Kirchner and Leopold Sögner

Abstract Questionnaire data is used to identify socio-demographic as well as the risk-awareness characteristics of users offended on the Internet. The data comprises a representative sample of 3,000 individuals, containing information on employment, education, age, the frequency of Internet usage and security measures taken by the users. Probit and logit regressions show that, conditional on using the Internet, being female and abstaining from using social media significantly reduces the risk to be offended on the Internet.

Susanne Kirchner
Department of Economics and Finance, Institute for Advanced Studies
Josefstädter Straße 39, 1080 Vienna, Austria

✉ kirchner@ihs.ac.at ,

Leopold Sögner
Institute for Advanced Studies and Vienna Graduate School of Finance

✉ soegner@ihs.ac.at

Archives of Data Science, Series A
(Online First)
KIT Scientific Publishing
Vol. 3, No. 1, 2018

DOI 10.5445/KSP/1000083488/02
ISSN 2363-9881

1 Introduction

The growing number of providers and users of Internet services as well as social communication networks also raises security issues and resulted in the emergence of cyber-crime research (see, e.g., Hartel et al, 2011). This article uses questionnaire data and applies statistical methods to identify users and their characteristics who were subject to some form of offense on the Internet and Social media by means of a cluster analysis. In a second step, we analyze how these characteristics are related to the likelihood of being offended on the Internet. In particular, we investigate how different safety measures taken by users affect their protection against cyber-crime.

In parallel to the emergence of cyber-crime science, for governmental institutions, such as the Ministry of the Interior and police authorities, the criminal aspects (such as theft of data, hacking, fraud, etc.) have become of particular interest (see, e.g., the study of Kirchner et al, 2015, in cooperation with the Austrian Federal Ministry of the Interior). To implement policies with the goal to improve cyber-security and to reduce crime (see, e.g., Becker, 1968; Freeman, 1999; Hartel et al, 2011; Dimkov, 2012), knowledge about the actual number of crimes committed, the socio-demographic structure of the users offended, factors (variables) raising the probability of an offense as well as the cost of Internet crime becomes important for a cost-benefit for hackers and the cost of cyber-crime (see, e.g., Kshetri, 2010; Anderson et al, 2013; Cook et al, 2014).

In addition, also governmental as well as non-governmental institutions provide guidelines how to responsibly use information technology. Such commandments are e.g. provided by the Computer Professionals for Social Responsibility (CPSR, 2015) or by Saferinternet.at (2016). For these institutions, knowledge on the socio-demographic structure of the users offended as well as user characteristics connected to offenses can be helpful to provide target group specific information, with the goal to increase the risk-awareness and to reduce the risk to be offended. Concerning the effectiveness of security awareness, Bullée et al (2015) showed in experiments that measures to increase security awareness turned out to be statistically significant.

Regarding academic publications in the field of cyber-crime research, almost recently, Hartel et al (2011) intensively searched through literature in various academic disciplines and concluded that “In spite of our efforts we have failed to find documented scientific studies of how Information Security effectively

prevents cyber-crime." By looking for causes of this gap, the authors claim that problems in information security are hardly reported to the police for several reasons. For example, a problem in the information security system need not result in crime.

To get more detailed information on the users to be offended and to quantify the risk to be offended on the Internet, this article uses the questionnaire data collected by Kirchner et al (2015) and identifies groups of offended Internet users. In particular, Section 2 discusses recent results obtained in cyber-crime research and presents the research questions. Then, Section 3 describes the data. To obtain information on the security-awareness and the socio-demographic characteristics of the users offended, Section 4 first presents results obtained by means of a cluster analysis. In a second step logit and probit regressions are performed to investigate the impact of user characteristics on the risk to be offended on the Internet. Section 5 concludes.

2 Cyber-Crime Research

Hartel et al (2011)[Section 2] define *crime science* as applying scientific methods to prevent and to detect disorder, particularly crime. Then, referring to Newman (2009), the authors define *cyber-crime* as "behaviour in which computers or networks are a tool, a target, or a place of criminal activity." By considering the historical development, "... cyber-crime emerged from hacking. Fraud schemes in relation with Social Engineering and other criminal activities were gradually added and connected to the technical and craft skills of the early hackers" (see Kochheim, 2016). While *information security research* is engaged in the development of software to increase IT security, *cyber-crime research* is connected to criminology and other social sciences with the goal to prevent cyber-crime (see, e.g., Hartel et al, 2011). For guidelines to perform information and communication technology research see e.g. Bailey et al (2012).

In addition, cyber-crime can be divided into "cyber-crime in a narrower sense", where offenses are committed by using the technologies of the Internet (e.g., illegal access to a computer system), and "cyber-crime in a broader sense" (see, e.g., Bundeskriminalamt, 2015, p. 17), where the Internet is used as communication medium for criminal activity (e.g., fraud, child pornography,

and the initiation of sexual contacts with minors). In this article we refer to the broader definition of cyber-crime.

Let us relate this article to recent literature in cyber-crime research: An overview on recent developments and results in cyber-crime science is e.g. provided in Hartel et al (2011) and Dimkov (2012).

Cyber-bullying was investigated in the empirical study of Hinduja and Patchin (2008). The authors used an online survey tool to collect data from 6,800 users in the time span December 2004 to January 2005. After focusing on the group of users not older than 17 years and data cleaning, the authors ended up with data from 1,378 users. The response variables constructed by the authors are two victimization variables ("general/serious cyber-bullying victimization") and two offending variables ("general/serious cyber-bullying offending"). Regarding serious cyber-bullying victimization, the authors observe (by applying logistic regression) that the time spent at the computer, school problems and being a bullying victim in real life are positively related to victimization. Other variables such as gender, age, black/white, and peer effects turned out to be insignificant. Due to the different age structure of the users, relating the study of Hinduja and Patchin (2008) to the results obtained in this article is difficult.

Information security awareness of Internet users was analyzed in Tsohou et al (2008) as well as Talib et al (2010). While Tsohou et al (2008) provide an overview on information security awareness, the study of Talib et al (2010) is based on survey data containing 333 observations. The authors argue that – compared to private use – at an individual's workplace clearer legislation and regulation about IT security exist. Because of this, the authors claim that learning about Internet security mainly takes place at an individual's workplace. Then, positive spill-over effects to security awareness at home are observed. Moreover, information on the "the perception of security in e-commerce B2C (business to customer) and C2C (costumer to customer) websites" is provided by Halaweh and Fidler (2008), who followed a qualitative approach by interviewing fifteen customers and twelve organizations' managers and their IT staff.

Kirchner and Angleitner (2016) analyzed which criminal-relevant phenomena and activities do occur in social media, to what extent did they reach so far, and which methods to attack users were applied. By using questionnaire data from Kirchner et al (2015) the study shows that Facebook (used by 62 % of the people asked in the questionnaire), WhatsApp (50 %) and YouTube (46 %) are those social media, which are used most frequently in the age group 14 - 49 years old.

Regarding police relevant issues, Kirchner and Angleitner (2016) observed that defective software/malware, hacking, fake accounts, cyber-mobbing (see also Schneider et al, 2013, and the literature cited there), phishing, cyber-bullying (see also Hinduja and Patchin, 2008), cyber-stalking, profile copying, sexting (see also Lee et al, 2013), and happy slapping are the most frequent ways how users were offended (the order of these terms corresponds to their frequency of occurrence).

3 Data and Research Questions

The research questions investigated in this article are: (i) "*What groups of persons show an insufficient problem-consciousness concerning cyber-crime and thus being at particular risk*?" and (ii) "*What variables increase/decrease the risk to be offended on the Internet*?". The first question will be investigated in Section 4.1 by means of a cluster analysis, while Section 4.2 applies binary regression techniques to analyze the risk to be offended on the Internet. The following section describes the data used to perform these analyses.

A very first step to investigate the risk of being offended on the Internet is to look on the number of notifications and complaints collected by police authorities. For example, the Austrian Ministry of the Interior collects the number of notifications on a yearly basis (for Austria, see e.g., BM.I, 2015, "Austrian Security Report"). This report shows the following: For 2014 a decline in the area of Internet crime is reported (−10.8 % compared to 2013), while for the last decade an increase from 1,794 notified offenses in 2005 to 8,966 notified offenses in 2014 is observed. After the significant rise in the last decade and the decrease in 2014, the criminal offenses are less than 10,000, which corresponds to approximately 0.1 % of the total Austrian population. The number of notified offenses is to be found mainly in the area of cyber-crime in a broader sense, and particularly, in the field of Internet fraud.

During the same periods, also the number of complaints increased enormously. In particular, from 1,151 in 2005 to 7,667 complaints in 2013. In parallel to the number of notifications, the complaints with respect to Internet fraud fell by 13.5 % in the year 2014. However, the value of 6,635 complaints in 2014, is imperceptibly higher than the value in 2012, where 6,598 complaints were observed. In addition, police authorities are also concerned about a large dark

field in the area of cyber-crime, and point out that new criminal phenomena are in progress (see Bundeskriminalamt, 2015).

To obtain more detailed information, this article uses data from the study of Kirchner et al (2015), where data on socio-demographic factors as well as on offenses on the Internet were collected for a target group of $N = 3,000$ representative users with an age between 14 and 49 years. More details are provided in Appendix 1. The $k = 21$ variables obtained from this study are:

y_n The binary variable *Attacked*, where 0 implies that the corresponding individual was not personally offended on the Internet or social media, while the variable is 1 if the user was offended personally. The variable y_n is derived and thereby defined as follows: Each individual n was asked whether she or he was already personally confronted with phishing, hacking, profile copying, fake accounts, malware, sexting, cyber stalking, cyber mobbing or cyber bullying. If at least one answer to these questions was "yes", we call individual n to be *offended* on the Internet. In formal terms the binary variable $y_n = 1$ if individual n was offended, while $y_n = 0$ if all answers were a "no". n denotes the index used for the corresponding person, $n = 1, \ldots, N$ (see question Q.1.$\cdot$ in Appendix 1). Hence, $O = \sum_{n=1}^{N} y_n$.

x_{n2} The variable *Frequency*, measuring the frequency of Internet and social network usage. This variable is an integer ranging from 0 to 2. A value of 0 denotes frequent use (category *a* in Question Q.2 Appendix 1), 1 stands for occasional use (category *b* in question Q.2) and 2 denotes no current use of social networks (categories *c* or *d* in question Q.2).

x_{n3} The binary variable *Gender*, where 0 stands for male and 1 for female (see question Q.3).

x_{n4} The integer variable *Age*, measured in years (see question Q.4).

x_{n5} The variable *Size of the City* approximates the number of inhabitants of the city where the individual currently lives. Here, the following categories are used: 1 abbreviates $< 10,000$ inhabitants, 2 stands for more than or equal to $10,000$ and less than $50,000$ inhabitants, 3 denotes more than or equal to $50,000$ and $< 100,000$ inhabitants, 4 denotes more than or equal to $100,000$ and $< 250,000$ inhabitants, while 5 stands for $\geq 250,000$ inhabitants (see question Q.5).

x_{n6} The integer variable *Employment* denotes the current employment status, where 0 stands for unemployment, 1 for part time employment and 2 for full employment. On leave, retirement, apprenticeship, civil- or military service and pupils are treated as missing value (see question Q.6 and the corresponding categories).

x_{n7} The variable *Human Capital* (Education), measuring the highest level of education obtained by individual n. This variable is equal to 1 if no school was completed, to 2 if the highest degree is from a secondary modern school ("Pflichschulabschluss" in the Austrian school system), to 3 if an apprenticeship, a school without general qualification for university entrance ("Berufsbildende mittlere Schule" or "Allgemeinbildende höhere Schule ohne Matura" in the Austrian school system) was completed, to 4 if a grammar school or an equivalent degree ("Berufsbildende höhere Schule" (e.g., HAK, HLW, HTL) in the Austrian school system) was completed, while 5 stands for some university degree (or (almost) equivalent degrees like "Abiturientenlehrgang, Kollege, Pädagogische Akademie" in the Austrian education system; see question Q.7 and the corresponding categories).

$x_{nS,j}$: Binary *Security/Incertitude* variables: The variable $x_{nS,j}$, $j = 1, \ldots, 14$, is set to 0 if an individual did not consider the corresponding security issues as relevant, while the value of the variable is one if the individual cared about that particular Internet security issue. These variables follow from the questions $Q.S1$ to $Q.S14$ provided in Table A-7.

Note that our definition of an offense also includes disruptions which need not be relevant for the police. However, since the questions raised in the questionnaire of Kirchner et al (2015) follow from expert interviews with the IT-division of the Austrian Ministry of the Interior, we consider the disruptions reported by the users either relevant or almost relevant for police authorities. Hence, the definition of an offense applied in this article can be considered to be slightly broader than a criminally relevant disruption. For the sample of $N = 3,000$, the number of people personally confronted with cyber-crime is $O = 470 \, (= \sum_{n=1}^{N} y_n)$. Comparing the rate $O/N \approx 16\,\%$ to the notification rate of approximately $0.13\,\%$, based on the data provided in BM.I (2015), strongly supports the arguments provided e.g. in Appendix A of Hartel et al (2011), who claimed that the number of offenses is above the number of offenses notified by the police. To

obtain an estimate of the notification rate, we use the Austrian population (≈ 8.5 million) in 2014 and an estimate of the percentage of users in Austria (82 %) in 2015 which is supposed to be a good approximation for the year 2014 (Statistik Austria, 2017). Then $8966/(0.82 \cdot 8500000) \approx 0.00129$. Note, that the difference between these two rates can either be caused by the claim that people do not contact the police, even if the offense is relevant for the police, or by the slightly broader definition of an offense used in this study ($y_n = 1$ can but need not imply that the offense is criminally relevant). Further reasons for differences are e.g. sampling effects or measurement errors. The differences observed between the male and the female population turned out to be small (this difference is also statistically insignificant at a 5 % significance level). In addition, we observed that from the $N = 3,000$ sample $N^* = 2,188$ individuals currently use social media and/or the Internet (i.e., the variable $x_{n2} \leq 1$). That is, $O/N^* \approx 21.48\,\%$ of the actual users were offended. In more formal terms, an estimate of the probability to use the Internet is $\widehat{\mathbb{P}}(x_{2n} \leq 1) = N^*/N = 0.7293$, while an estimate of the conditional probability $\mathbb{P}(y_n = 1 | x_{2n} \leq 1)$ is $\widehat{\mathbb{P}}(y_n = 1 | x_{2n} \leq 1) = 0.2148$.

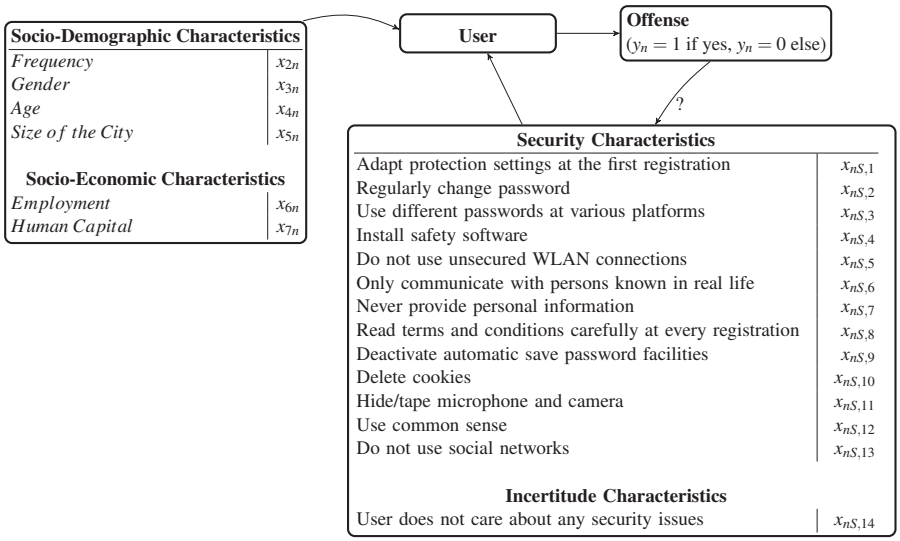

Figure 1: Variables and Effects This figure plots the socio-demographic, the socio-economic and the security/incertitude variables obtained from the questionnaire study of Kirchner et al (2015). Their interrelationships are denoted by $\rightarrow$, the observation index is $n = 1, \ldots, N = 3,000$. The last columns in the corresponding boxes present the abbreviations used for the corresponding variables.

The data, abbreviated by $\mathbf{X} \in \mathbb{R}^{N\times k}$, contains the observations $\mathbf{x}_n = (y_n, x_{n2}, \ldots, x_{nk})^\top \in \mathbb{R}^k$ for $n = 1,\ldots,N = 3{,}000$ and $k = 1,\ldots,21$. For vectors and matrices boldface notation is applied. That is to say, $\mathbf{z} \in \mathbb{R}^p$ denotes a p-dimensional column vector, $\mathbf{Z} \in \mathbb{R}^{p\times q}$ a $p \times q$ matrix. z_i denotes the i-th coordinate of the column vector $\mathbf{z}$. $\mathbf{z}^\top$ denotes the transpose of $\mathbf{z}$. $\mathbb{I}(\cdot)$ abbreviates an indicator function. Estimates are labeled by the superscript $\widehat{\ }$. Sample means and sample standard deviations for the variables y_n, x_{ni}, $i = 2,\ldots,7$, and $x_{nS,j}$ are provided in the last column of Table 1. If no answer is provided or if the answer "don't know" is chosen for some variable by individual n, we obtain a missing value. For y_n, x_{n2} and x_{n3} no missing values are observed. For the variables age, inhabitants and human capital two, thirty and eighteen missing values are observed. For the variable *Employment* where on leave, retirement, apprenticeship, civil- or military service and pupils are treated as missing values we get 616 missing values, while for each of the security/incertitude variables $x_{nS,j}$ 681 missing values are observed. For the cluster analysis performed in Section 4 all $N = 3{,}000$ observations can be used by setting the contribution for the corresponding variable to zero when obtaining the distance function, while for the regression analysis observations containing missing values were excluded by the software package.

The variables presented in Figure 1, contain the socio-demographic characteristics *Frequency*, *Gender*, *Age* and the *Size of the City*, the socio-economic characteristics, *Employment* and *Human Capital* (education) as well as some security/incertitude variables. In the cluster analysis performed in Section 4.1 we shall investigate similarities between a subset of these variables, while the regression results presented in Section 4.2 will analyze how these characteristics influence the probability to be offended. We expect the probability to be offended to decrease in the variable *Frequency* (note that $x_{n2} = 0$ denotes frequent use, ..., $x_{n2} = 2$ currently no use). From the previous section we expect no significant effects for the variable *Gender* based on the results of Hinduja and Patchin (2008). Although Hinduja and Patchin (2008) used a different age group, we also do not expect significant effects for the variable *Age*. Based on our literature review the effects of the variables *Size of the City* and *Human Capital* are unclear. By applying the argument of Tsohou et al (2008) that "security awareness increases due to clearer regulation at work", we expect the variable *Employment* to decrease the probability to be offended. For all security variables $x_{nS,1}$ to $x_{nS,13}$ we expect the probability of an offense to diminish, while the opposite effect is expected for the incertitude variable $x_{nS,14}$.

4 Results

4.1 Groups of Offended and Non-offended Users

The goal of the following exploratory analysis is to find out which individuals share similar characteristics. In particular, we want to find out whether there are differences (i) between offended and non-offended individuals or (ii) within the subset of offended users. To do this we cluster the data described in Section 3, such that the individuals in the same cluster have stronger similarities than the individuals collected in the other clusters. In particular, the agglomerate hierarchical clustering algorithm `agnes` described in Kaufman and Rousseeuw (1990, Chapter 5) is applied. The observations used to perform the cluster analysis are $\mathbf{x}_n = \left(y_n, x_{n2}, \ldots, x_{n7}, x_{nS,1}, \ldots, x_{nS,14}\right)^\top \in \mathbb{R}^k$, where $k = 21$, for $n = 1, \ldots, N = 3000$. The data used to perform the cluster analysis is abbreviated by $\mathbf{X} \in \mathbb{R}^{N \times k}$, collecting the observations $\mathbf{x}_n$, $n = 1, \ldots, N$. Additionally, a distance function measuring the dissimilarity between the observations $\mathbf{x}_n$ and $\mathbf{x}_m$ has to be chosen. Since the data $\mathbf{X}$ contains the numerical variable *Age*, the binary variables *Offense*, *Gender* and $x_{nS,j}$ as well as variables *Frequency*, *Size of the City*, *Employment* and *Human Capital* measured on an ordinal scale, we follow suggestions in the literature and apply Gower-distances (1971). *Gower-distances* are also implemented in the software package `R` as described in Maechler et al (2015). In addition to Gower-distances we also performed a cluster analysis with L_1 (= sum of absolute distances or Manhattan distances) and Euclidean distances. The differences to the results obtained with Gower-distances are relatively small. In addition, we performed robustness checks such as excluding the variable *Gender*, only working with one $x_{nS,j}$, etc. Although the choice of the variables impacts the output obtained by the cluster analysis, the result that for a group of offended users we observe in parallel a group of non-offended users but with the other characteristics quite similar remains stable.

By applying this clustering technique to our data $\mathbf{X}$, we observe a high agglomerative coefficient of $AC \approx 0.98$, measuring the quality of the clustering method applied to the data (see, e.g., Kaufman and Rousseeuw, 1990, p. 211). Based on the dendrogram and with the goal to get a parsimonious description of the data, we decided to present the result where the data $\mathbf{X}$ is clustered into twelve groups. This decision is based on the observation that for the branches

on the top of the clustering tree larger differences are observed, while for a larger number of clusters the differences in the variables of interest for this study become small.

Table 1 presents results when $I = 12$ groups are considered. The columns 2 to 13 present the group-specific mean values and the group-specific standard deviations within the corresponding cluster $\mathscr{C}_i$. The last column presents the sample means and the sample standard deviations for each variable, obtained from $N = 3,000$ observations. The last row presents the number of individuals assigned to cluster $\mathscr{C}_i$, $i = 1,\ldots,I = 12$. Note that the mean value for the variable $Attack$ corresponds to the percentage of the individuals offended on the net, i.e. $\frac{O}{N} = \frac{470}{3000} = 0.1567$.

From Table 1 we observe that all individuals offended on the Internet are contained in the clusters $\mathscr{C}_2$, $\mathscr{C}_4$, $\mathscr{C}_6$, $\mathscr{C}_7$ and $\mathscr{C}_{12}$, where $\mathscr{C}_4$ and $\mathscr{C}_7$ contain offended users only. The bulk of offended users is contained in Class $\mathscr{C}_4$ (this class contains 457 of 470 offended individuals). In for this class relatively small deviations to the sample means (provided in the last column) are observed. Excluding offenses, the characteristics of the individuals in $\mathscr{C}_4$ are also very similar to the individuals in $\mathscr{C}_2$ (where only a very small number of offended users is included). $\mathscr{C}_7$ contains mainly young, male users with a slightly higher mean level of employment. In this group a lot but not all security awareness variables are high.

Summing up, regarding the security awareness measured by the variables $x_{nS,j}$, $j = 1,\ldots,14$, the clustering results provided in Table 1 do not show very clear results (i.e., to distinguish offended from non-offended users based on $x_{nS,j}$). Hence, the impact of $x_{nS,j} = 1$, $j = 1,\ldots,14$, on the probability to be offended in the Internet seems to be low. In addition, we do not observe strong differences between the offended and the non-offended clusters regrading the variables *Gender*, *Employment*, *Human Capital*, *Size of the City* and *Age*. That is, for a cluster of offended users we find in parallel a cluster of non-offended users where the socio-economic characteristics are quite similar. Therefore, based on the cluster analysis we do not expect strong impacts of these variables on the probability to be offended. To obtain more detailed results we proceed to estimate logit and probit models in the following Section 4.2.

Table 1: Results obtained from the Cluster Analysis. Results obtained from the cluster analysis. Data set $\mathbf{X}$, $N = 3,000$ observations, $k = 21$ variables, $I = 12$ clusters and Gower-distances. For each variable the first row presents group-specific sample means in the corresponding cluster $\mathscr{C}_i$, $i = 1, \ldots, 12$, while the second row presents the group-specific sample standard deviations. The last column presents the mean values and the sample standard deviations (SD) for the corresponding variables, obtained from all observations $n = 1, \ldots, N$. The last row presents the number of individuals assigned to cluster $\mathscr{C}_i$. *NA* denotes "not available".

	Cluster $\mathscr{C}_i$												
Variable	1	**2**	3	**4**	5	**6**	**7**	8	9	10	11	**12**	mean/SD
Attacked	0.00	0.00	0.00	1.00	0.00	0.16	1.00	0.00	0.00	0.00	0.00	0.30	0.157
	0.00	0.03	0.00	0.00	0.00	0.37	0.00	0.00	0.00	0.00	0.00	0.48	0.364
Frequency	0.10	1.37	0.03	1.32	0.35	1.26	1.33	1.80	1.50	0.34	0.33	1.10	0.957
	0.31	0.50	0.16	0.65	0.55	0.79	0.58	0.42	0.71	0.61	0.58	0.57	0.763
Gender	0.41	0.51	0.53	0.47	0.29	0.26	0.33	0.50	1.00	0.28	0.00	0.30	0.497
	0.50	0.50	0.50	0.50	0.46	0.45	0.58	0.53	0.00	0.45	0.00	0.48	0.500
Age	21.67	31.97	40.55	32.98	35.65	28.13	29.67	35.50	33.00	27.72	33.00	38.60	34.229
	4.89	9.91	7.21	9.70	10.49	12.19	6.51	7.18	4.24	7.17	2.65	8.14	10.048
Inhabitants	2.52	2.45	1.95	2.57	3.60	2.22	2.67	3.10	5.00	1.69	5.00	3.40	2.360
	1.59	1.74	1.54	1.79	1.81	1.69	2.08	2.02	0.00	1.20	0.00	2.07	1.720
Employment	1.36	1.21	1.29	1.12	0.13	1.03	1.67	1.00	2.00	1.07	2.00	1.00	1.197
	0.49	0.47	0.46	0.49	0.34	0.42	0.58	0.47	0.00	0.37	NA	0.50	0.496
Human Capital	2.82	3.77	3.79	3.88	3.66	3.13	4.33	3.90	4.50	3.41	4.00	4.00	3.762
	0.91	1.04	0.99	0.97	0.99	1.21	0.58	0.88	0.71	0.73	0.00	0.87	1.020
$x_{nS,1}$	0.56	0.72	0.53	0.67	0.54	0.00	1.00	0.60	0.50	0.00	0.50	0.90	0.674
	0.51	0.45	0.50	0.47	0.51	0.00	0.00	0.52	0.71	0.00	0.71	0.32	0.469
$x_{nS,2}$	0.20	0.43	0.35	0.46	0.40	0.00	0.33	0.40	1.00	0.12	0.50	0.40	0.420
	0.41	0.50	0.48	0.50	0.50	0.00	0.58	0.52	0.00	0.33	0.71	0.52	0.494
$x_{nS,3}$	0.72	0.65	0.46	0.67	0.46	0.00	1.00	0.80	1.00	0.06	0.50	1.00	0.627
	0.46	0.48	0.50	0.47	0.51	0.00	0.00	0.42	0.00	0.24	0.71	0.00	0.484
$x_{nS,4}$	0.60	0.76	0.60	0.78	0.60	0.00	1.00	0.50	0.00	0.00	0.50	1.00	0.731
	0.50	0.42	0.49	0.42	0.50	0.00	0.00	0.53	0.00	0.00	0.71	0.00	0.364
$x_{nS,5}$	0.32	0.52	0.47	0.53	0.40	0.00	0.33	0.50	0.50	0.00	0.50	0.90	0.507
	0.48	0.50	0.50	0.50	0.50	0.00	0.58	0.53	0.71	0.00	0.71	0.32	0.500
$x_{nS,6}$	0.40	0.77	0.67	0.72	0.69	0.00	1.00	0.60	0.00	0.18	0.00	1.00	0.726
	0.50	0.42	0.47	0.45	0.47	0.00	0.00	0.52	0.00	0.39	0.00	0.00	0.446
$x_{nS,7}$	0.48	0.74	0.68	0.74	0.57	0.00	0.33	0.50	1.00	0.18	1.00	0.90	0.716
	0.51	0.44	0.47	0.44	0.50	0.00	0.58	0.53	0.00	0.39	0.00	0.32	0.451
$x_{nS,8}$	0.24	0.33	0.36	0.37	0.29	0.00	0.00	0.50	0.50	0.00	0.50	0.50	0.333
	0.44	0.47	0.48	0.48	0.46	0.00	0.00	0.53	0.71	0.00	0.71	0.53	0.472
$x_{nS,9}$	0.60	0.57	0.51	0.58	0.34	0.00	0.67	0.50	1.00	0.00	0.50	0.70	0.552
	0.50	0.50	0.50	0.49	0.48	0.00	0.58	0.53	0.00	0.00	0.71	0.48	0.497
$x_{nS,10}$	0.64	0.59	0.51	0.63	0.57	0.00	1.00	0.60	0.50	0.00	1.00	0.60	0.583
	0.49	0.49	0.50	0.48	0.50	0.00	0.00	0.52	0.71	0.00	0.00	0.52	0.493
$x_{nS,11}$	0.40	0.23	0.19	0.22	0.11	0.00	0.00	0.10	1.00	0.00	0.50	0.20	0.443
	0.50	0.42	0.40	0.42	0.32	0.00	0.00	0.32	0.00	0.00	0.71	0.42	0.221
$x_{nS,12}$	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	0.006
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.080
$x_{nS,13}$	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.016
	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.127
$x_{nS,14}$	0.00	0.00	0.05	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.020
	0.00	0.00	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.139
Members	49.00	1549.00	788.00	457.00	62.00	38.00	3.00	10.00	2.00	29.00	3.00	10.00	3000.000

4.2 The Risk to be Offended on the Internet

In this section we investigate the second question, that is, we consider variables increasing or decreasing the risk to be offended on the Internet. In this analysis we consider the $N^* = 2,188$ subsample consisting of individuals currently using the Internet. Hence, all the following results are conditional on using the Internet. We implicitly assume that only people actually using the Internet were subject to an offense. Hence, $\mathbb{P}(y_n = 1|\tilde{\mathbf{x}}_n,\ x_{n2} = 2) = \mathbb{P}(y_n = 1|\ x_{n2} = 2) = 0$. To investigate this question we have to account for the fact that y_n is a binary variable. In formal terms we consider the events $\{y_n = 1\}$ and $\{y_n = 0\}$, where users participate, i.e. $\{x_{n2} \leq 1\}$. Logit and probit regressions (see, e.g., Greene, 1997; Cameron and Trivedi, 2005) are applied to obtain estimates how the conditional probability $\mathbb{P}(y_n = 1|\tilde{\mathbf{x}}_n,\ x_{n2} \leq 1)$ depends on the explanatory variables $\tilde{\mathbf{x}}_n$. To avoid strong assumptions on the possible effects arising from the variables measured on an ordinal scale, we define the binary (*dummy*) variable $\mathbb{I}_{n2}(x_{n2} = 1)$ which is equal to 1 if the value of the variable frequency is $x_{n2} = 1$, otherwise this variable is zero. In a similar way we obtain the binary variables $\mathbb{I}_{n5}(x_{n5} = 1)$ (living in small cities), $\mathbb{I}_{n5}(x_{n5} \geq 3)$ (living in lage(r) cities), $\mathbb{I}_{n6}(x_{n6} = 1)$ (part time employment), $\mathbb{I}_{n6}(x_{n6} = 2)$ (full employment), $\mathbb{I}_{n7}(x_{n7} = 3)$ (medium level of education), $\mathbb{I}_{n7}(x_{n7} = 4)$ (high school education) and $\mathbb{I}_{n7}(x_{n7} = 5)$ (university education). By means of a 1 as the first coordinate of the vector of explanatory variables $\tilde{\mathbf{x}}_n$, we include an intercept term. Hence, our explanatory variables are $\tilde{\mathbf{x}}_n := (1, \mathbb{I}_{n2}(x_{n2} = 1), x_{n3}, x_{n4}, \mathbb{I}_{n5}(x_{n5} = 1), \mathbb{I}_{n5}(x_{n5} \geq 3), \mathbb{I}_{n6}(x_{n6} = 1), \mathbb{I}_{n6}(x_{n6} = 2), \mathbb{I}_{n7}(x_{n7} = 3), \mathbb{I}_{n7}(x_{n7} = 4), \mathbb{I}_{n7}(x_{n7} = 5), x_{nS,1}, \ldots, x_{nS,14})^\top \in \mathbb{R}^{k^*}$, such that the number of exploratory variables is $k^* = 25$.

The relatively high number of explanatory variables k^* could result in (almost) collinear regressors. When including the human capital variable $\mathbb{I}_{n7}(x_{n7} = 2)$ we observe almost collinear variables. Therefore, $\mathbb{I}_{n7}(x_{n7} = 2)$ was excluded and our results regarding human capital are measured against low or very low education. In addition, we also checked the correlations of the $k^* = 25$ regressors. This and missing warnings on collinearity provided by the `R` package did not indicate problems related to multicollinearity. In addition, we abstract from feedback effects from $\tilde{\mathbf{x}}_n$ on y_n (in more technical terms we assume that the regressors $\tilde{\mathbf{x}}_n$ are exogenous; see, e.g., Davidson and MacKinnon, 1993, p. 624-627). With probit and logit models $\mathbb{P}(y_n = 1|\tilde{\mathbf{x}}_n,\ x_{n2} \leq 1) =$

$\mathbb{E}(y_n = 1|\tilde{\mathbf{x}}_n,\ x_{n2} \leq 1) = F(\boldsymbol{\beta}^\top \tilde{\mathbf{x}}_n)$, where $\boldsymbol{\beta} \in \mathbb{R}^{k^*}$. The regression parameter β_i describes the impact of $\tilde{x}_{ni}$, i.e. the ith coordinate of $\tilde{\mathbf{x}}_n$, on the conditional probability $\mathbb{P}(y_n = 1|\tilde{\mathbf{x}}_n,\ x_{n2} \leq 1)$ [equal to the conditional expectation $\mathbb{E}(y_n = 1|\tilde{\mathbf{x}}_n,\quad x_{n2} \leq 1)$]. To assign "self-explaining subscripts" to the regression parameters β_i we define the set $\mathbb{I}^*$, which contains the $k^* = 25$ names of explanatory variable provided in the first column of Table 2. That is, $\mathbb{I}^* := \{$*Intercept*, *Frequency* $\mathbb{I}_{n2}(x_{n2} = 1), \ldots,$ *Human Capital* $\mathbb{I}_{n7}(x_{n7} = 5)$, *Security* $x_{nS,1}, \ldots,$ *Security* $x_{nS,13}$, *Incertitude* $x_{nS,14}\}$. Then, the notation β_i, $i = 1, \ldots, k^*$, is equivalent to $\beta_{Intercept}$, $\beta_{Frequency\ \mathbb{I}_{n2}(x_{n2}=1)}$, $\beta_{Gender\ x_{n3}}$, $\ldots$, $\beta_{Incertitude\ x_{nS,14}}$. The function $F(\cdot)$ is called link function. For the logit model the link function is provided by the logistic function, i.e.

$$\begin{aligned} \mathbb{P}(y_n = 1|\tilde{\mathbf{x}}_n,\ x_{n2} \leq 1) &= \frac{e^{\boldsymbol{\beta}^\top \tilde{\mathbf{x}}_n}}{1 + e^{\boldsymbol{\beta}^\top \tilde{\mathbf{x}}_n}}\ , \quad \text{while for the probit model} \\ \mathbb{P}(y_n = 1|\tilde{\mathbf{x}}_n,\ x_{n2} \leq 1) &= \Phi\left(\boldsymbol{\beta}^\top \tilde{\mathbf{x}}_n\right)\ , \end{aligned} \tag{1}$$

where $\Phi(\cdot)$ abbreviates the distribution function of the standard normal distribution. In this article parameter estimates, denoted by $\widehat{\boldsymbol{\beta}}$, of the parameter vector $\boldsymbol{\beta}$ are obtained by means of maximum likelihood estimation (by using the `glm` function contained in the `R` package `AER`). To investigate the question how $\tilde{x}_{ni}$ affects $\mathbb{P}(y_n = 1|\tilde{\mathbf{x}}_n)$, the marginal effects

$$\frac{\partial}{\partial \tilde{x}_{ni}} \mathbb{E}(y_n = 1|\tilde{\mathbf{x}}_n,\ x_{n2} \leq 1) = F(\boldsymbol{\beta}^\top \tilde{\mathbf{x}}_n)\beta_i\ , \tag{2}$$

can be obtained for any $i = 1, \ldots, k^*$ (see, e.g., Greene, 1997; Cameron and Trivedi, 2005). In contrast to the linear regression model, the marginal effects described in (2) depend on the value of $\tilde{\mathbf{x}}_n$ where (2) is evaluated. In the following analysis, the term ME_i abbreviates the marginal effect $\frac{\partial}{\partial \tilde{x}_{ni}} \mathbb{E}(y_n = 1|\tilde{\mathbf{x}}_n,\ x_{n2} \leq 1)$ evaluated at the conditionally expected value $\mathbb{E}(\tilde{\mathbf{x}}_n|x_{n2} \leq 1)$. We obtain an estimate of the marginal effect, $\widehat{ME}_i$, by replacing $\boldsymbol{\beta}$ and $\mathbb{E}(\tilde{\mathbf{x}}_n|x_{n2} \leq 1)$ by their finite sample analogs $\widehat{\boldsymbol{\beta}}$ and $\bar{\tilde{\mathbf{x}}}_n = \frac{1}{N^*} \sum_{n=1}^{N^*} \tilde{\mathbf{x}}_n$.

In contrast to the assumption of exogenous regressors, some users might have decided to "install safety software", to "read terms and conditions carefully at every registration", etc. *after* they had been offended and *before* they had been interviewed (in which case regressor endogeneity arises). If this is the case, we also obtain an impact of the offense variable y_n to the security/incertitude variables $x_{nS,\cdot}$, marked by $\overset{?}{\rightarrow}$ in Figure 1, or to x_{n2} (for example a user stopped

to use the Internet, i.e. $x_{n2} = 2$, *after* she or he was offended). In this case we would observe reverse causality. If there are serious concerns that the persons interviewed behaved in this way, instrumental variable estimation should be performed, where we claim that finding good instruments for the given regression is a difficult problem. Based on these arguments, we suggest to include questions on possible changes in the users' behavior when further questionnaire studies are performed.

Table 2: Results obtained from the Logit and Probit Regression. Results obtained from the logit and probit regressions. $N^* = 2,188$ observations. y_n, i.e. "personally offended", is the binary dependent variable, while the explanatory variables are provided in the first column (an intercept term is included). The second and the sixth column provide the maximum likelihood estimates $\widehat{\beta}_i$, $i = 1, \ldots, k^*$, the third and the seventh column provide standard errors, while the forth and the eighth column provide p-values. Significant variables at the 5 % significance level in boldface. The fifth and the last column show estimates of the marginal effects ME_i, $i = 1, \ldots, k^*$.

	Logit Model				Probit Model			
Variable	$\widehat{\beta}_i$	SE_i	p-value	$\widehat{ME}_i$	$\widehat{\beta}_i$	SE_i	p-value	$\widehat{ME}_i$
Intercept	**-1.5167**	0.6538	0.0203	-0.2042	**-0.9356**	0.3648	0.0103	-0.2388
Frequency $\mathbb{I}_{n2}(x_{n2} = 1)$	-0.2690	0.2064	0.1924	-0.0362	-0.1414	0.1186	0.2330	-0.0361
Gender x_{n3}	**-0.4478**	0.1803	0.0130	-0.0603	**-0.2535**	0.1010	0.0121	-0.0647
Age x_{n4}	-0.0038	0.0094	0.6872	-0.0005	-0.0016	0.0053	0.7634	-0.0004
Inhabitants $\mathbb{I}_{n5}(x_{n5} = 1)$	0.1407	0.2364	0.5517	0.0189	0.0815	0.1318	0.5365	0.0208
Inhabitants $\mathbb{I}_{n5}(x_{n5} \geq 4)$	0.2008	0.2500	0.4219	0.0270	0.1195	0.1400	0.3935	0.0305
Employment $\mathbb{I}_{n6}(x_{n6} = 1)$	-0.5126	0.3489	0.1418	-0.0690	-0.2974	0.2050	0.1468	-0.0759
Employment $\mathbb{I}_{n6}(x_{n6} = 2)$	-0.3927	0.3785	0.2996	-0.0529	-0.2220	0.2208	0.3146	-0.0567
Human Capital $\mathbb{I}_{n7}(x_{n7} = 3)$	0.2572	0.4142	0.5347	0.0346	0.1532	0.2224	0.4910	0.0391
Human Capital $\mathbb{I}_{n7}(x_{n7} = 4)$	0.7279	0.4201	0.0832	0.0980	0.4076	0.2275	0.0731	0.1040
Human Capital $\mathbb{I}_{n7}(x_{n7} = 5)$	0.5718	0.4120	0.1651	0.0770	0.3217	0.2221	0.1474	0.0821
Security $x_{nS,1}$	-0.1711	0.1943	0.3787	-0.0230	-0.0898	0.1103	0.4154	-0.0229
Security $x_{nS,2}$	-0.2662	0.1741	0.1262	-0.0358	-0.1481	0.0981	0.1312	-0.0378
Security $x_{nS,3}$	0.3485	0.1916	0.0690	0.0469	0.1930	0.1071	0.0715	0.0492
Security $x_{nS,4}$	0.4077	0.2124	0.0549	0.0549	0.2242	0.1171	0.0555	0.0572
Security $x_{nS,5}$	-0.0252	0.1784	0.8877	-0.0034	-0.0063	0.1013	0.9505	-0.0016
Security $x_{nS,6}$	-0.2882	0.2005	0.1506	-0.0388	-0.1649	0.1149	0.1512	-0.0421
Security $x_{nS,7}$	0.3695	0.2164	0.0878	0.0497	0.2205	0.1203	0.0668	0.0563
Security $x_{nS,8}$	**0.3405**	0.1714	0.0471	0.0458	0.1829	0.0974	0.0603	0.0467
Security $x_{nS,9}$	0.0247	0.1793	0.8906	0.0033	0.0099	0.1015	0.9222	0.0025
Security $x_{nS,10}$	-0.0787	0.1783	0.6589	-0.0106	-0.0483	0.1006	0.6313	-0.0123
Security $x_{nS,11}$	-0.2041	0.2075	0.3253	-0.0275	-0.1202	0.1159	0.2994	-0.0307
Security $x_{nS,12}$	1.4315	1.0335	0.1660	0.1927	0.8700	0.6421	0.1755	0.2220
Security $x_{nS,13}$	**-2.0427**	1.0314	0.0476	-0.2750	**-0.9950**	0.4496	0.0269	-0.2539
Incertitude $x_{nS,14}$	0.2829	0.6232	0.6499	0.0381	0.1666	0.3507	0.6348	0.0425

Table 2 provides the regression results. By looking at the probability values (*p-values*), we observe that the regression intercept and the variable *Gender* are highly statistically significant for both models. Since the estimates of $\beta_{Gender\ x_{n3}}$ are negative, a smaller risk of an offense on the Internet is observed for women. By means of the marginal effects we observe that a rise in the variable *Gender* by an infinitesimal unit, decreases the probability to be offended by approximately 6 % times this infinitesimal unit. When applying a significance level of 5 % the variable "*Frequency* $\mathbb{I}_{n2}(x_{n6} = 2)$" and the employment dummies are statistically insignificant. At the 15 % significance level the variable $\mathbb{I}_{n6}(x_{n6} = 1)$ (i.e., non-full time employment) is significant. Since higher employment reduces the risk to be offended on a significance of 15 %, the regressions provide very weak support for the learning arguments provided Talib et al (2010). Higher education, measured by the variable *Human Capital*, interestingly hardly changes the probability to be offended. Only for an education level corresponding to high school (described by the dummy variable $\mathbb{I}_{n7}(x_{n7} = 4)$) an increase in the probability to be offended is observed at a significance level close to 10 %. The impacts of the variables *Age*, and *Size of the City* are statistically insignificant (when applying significance levels ≤ 10 %). Finally, we investigate the impacts arising from the various $Security$ variables $x_{nS,j}$. For both the logit and the probit model, only the variables $x_{nS,3}$ ("use different passwords at various platforms"), $x_{nS,4}$ ("install safety software"), $x_{nS,7}$ ("never provide personal"), $x_{nS,8}$ ("read terms and conditions carefully at every registration") and $x_{nS,13}$ ("do not use social networks") are significant at the 10 % level, while only $x_{nS,13}$, for both models, and $x_{nS,8}$, for the logit model only, are significant also at a 5 % significance level. The sign of $\widehat{\beta}_{\text{Secturity } x_{nS,13}}$ is negative as expected (i.e. the probability of an offense decreases), while – in contrast to our expectations – the estimates $\widehat{\beta}_{\text{Security } x_{nS,3}}$, $\widehat{\beta}_{\text{Security } x_{nS,4}}$, $\widehat{\beta}_{\text{Secturity } x_{nS,7}}$ and $\widehat{\beta}_{\text{Secturity } x_{nS,8}}$ have positive signs. Although the probability values for these parameters are larger than 5 %, as already discussed before, some of the responses/security activity measured by questionnaire might happened after an offense and a reverse causality problem is observed.

Summing up, by using the results of both regression methods and a significance level of 5 %, only the variables "x_{n3} – gender: being female" and "$x_{nS,13}$ – do not use social networks" turned out to be statistically significant and reduce the probability to be offended on the Internet.

5 Conclusions

To prevent and reduce the risk of individuals to be offended on the Internet, more detailed information on the socio-demographic as well as the risk-awareness characteristics of the users with respect to Internet security becomes necessary. This study uses questionnaire data from 3,000 Austrian individuals, recently collected by Kirchner et al (2015), to provide information on these issues. The sample used in this article, contains information on employment, education, age and the frequency of Internet usage.

First, by means of a cluster analysis we investigate the question regarding the groups of persons being offended on the Internet. The cluster analysis does not provide a very clear picture. For each cluster containing offended users, we observe – in parallel – a cluster with non-offended users having similar characteristics. Second, we analyze the question whether the characteristics of the users such as age and gender as well as various protection methods applied by the users increase or decrease the risk to be offended on the Internet. By means of probit and logit regressions we observe that being female and to abstain from using social media significantly reduces the risk to be offended on the Internet.

Acknowledgements The authors thank B. Angleitner, M. Gstrein, U. Röhsner (MAKAM Research), M. Popolari and A. Mattern (both Austrian Federal Ministry of the Interior; Sektion IV (Department IV/6)) as well as especially Robert Kunst, Andreas Geyer-Schulz and an anonymous referee for interesting discussions and comments. We gratefully acknowledge funding from the Austrian KIRAS program (KIRAS security research program) financed by the Austrian Federal Ministry for Transport, Innovation and Technology.

Appendix

Further Information about the Data

The study of Kirchner et al (2015) is based on two surveys: The first sample comprises data from the Austrian population with an age between 14 and 49 years. The second sample considers parents (both or one parent) of children aged 10 to 13 years. In order to create the basis for the surveys and focus groups, 8 interviews with experts of the IT-division of the Austrian Ministry of the Interior (BM.I) as well as police-attorneys have been conducted. During the expert-interviews the problems of using the social media and future challenges were discussed. The results of the expert-interviews were used to design the questionnaires.

To obtain these data, Computer Assisted Telephone Interviews were performed. The data finally consists of 3,000 Austrians aged 14 to 49 years and 500 parents of children aged 10 to 13 years by using a standardized questionnaire. According to the requirements of the study, the characteristics of gender, age and place of residence (federal state) were considered as representative criteria. To obtain these data, in total, about 50,000 people were contacted in order to achieve the desired 3,500 interviews. This corresponds to a response rate of around 7 %. For about 37 % of the calls, no one picked up; about 18 %, the number from the phone book was invalid. Approximately 22 % refused to participate in the survey and approximately 4 % broke off the interview during the conversation.

The $N = 3,000$ survey was held in the period from July 9, 2014 to October 12, 2014. With the goal to obtain information on young users, in addition to the $N = 3,000$ sample used in this article, Kirchner et al (2015) interviewed 500 parent(s) from December 11, 2014 until May 1, 2015. In those cases where the parents had more than one child in this age group, they were asked at the beginning of the interview how many children in this age group they have - and a random selection was set to which of their children they should refer.

In the statistical analysis provided in this article only the $N = 3,000$ sample is used. For this sample of $N = 3,000$ interviews we observe the following: Let ζ abbreviate some attribute of the population measured in percentage terms. Then, given some point estimate $\widehat{\zeta}$ based on the sample $\mathbf{X}$ of size $N = 3,000$, the 95 % confidence interval (based on the normal approximation following from the asymptotic analysis) is $\left[\widehat{\zeta} - 1.8\,\%, \widehat{\zeta} + 1.8\,\%\right]$. In addition, by comparing

the percentages observed for the population to their sample analogs, we observe that all percentages observed for the population are contained in the interval "value observed in the sample ± standard error". By this we consider the survey samples as representative. That is, the distribution of the characteristics of gender, age and place of residence in the sample corresponds to that in the population. The subset of questions of the study of Kirchner et al (2015) used in this article are provided in the Tables A-3 to A-7. In particular, Table A-3 provides the questions used to construct the binary variable y_n, which is equal to one if user n was offended and zero else. Tables A-5 and A-6 provide the questions used to construct the socio-economic and the socio-demographic variables x_{n2} to x_{n7}. Finally, Table A-7 shows the list of questions to construct the security and incertitude variables $x_{nS,j}$, $j = 1, \ldots, 14$. Since responses to the open questions were very rare, the answers to these questions cannot be exploited in our analysis.

Table A-3: Questions related to the variable offense, y_n (1/2).

Questions in German	Questions in English
Opferwerdung	**Victimization**
Waren Sie selbst oder jemand aus Ihrem Umfeld bereits einmal von einer der folgenden kriminellen Aktivitäten auf Facebook oder einem anderen sozialen Netzwerk betroffen?	Have you or someone in your area already been affected by one of the following criminal activities on Facebook or any other social network?
a) Ja, ich selbst, und zwar durch:	a) Yes, I myself, by means of:
Q.1.1 *Phishing* – Abfrage von Benutzerdaten über gefälschte Anmeldeseiten oder gefälschte E-Mails von Facebook oder einem anderen sozialen Netzwerk.	*Phishing* – Fake log-in pages or fake e-mails from Facebook or another social network (user data are requested).
Q.1.2 *Hacking* - Illegale Einsicht eines Accountes, um z. B. Passwörter oder Kreditkartendaten zu stehlen.	*Hacking* – An account is illegally viewed, e.g. to steal passwords or credit cards.
Q.1.3 *Profile Copying/-Cloning / Identitätsdiebstahl* – Jemand anderes nimmt in Facebook oder in einem anderen sozialen Netzwerk eine fremde Identität an.	*Profile Copying / Cloning* – Someone else uses a foreign identity in Facebook or another social network.

Table A-4: Questions related to the variable offense, y_n (2/2).

Questions in German	Questions in English
Q.1 **Opferwerdung**	**Victimization**
Q.1.4 *Fake-Accounts* – Anlegen gefälschter Accounts in Facebook oder in einem anderen sozialen Netzwer, um zu anderen Freunden bzw. auf deren Profile zu gelangen.	*Fake Accounts* – Using fake accounts in Facebook or any other social network to access to other friends data or to access their profiles.
Q.1.5 *Schadhafte Software / Malware* – Angriff mit Computerviren oder Trojanern, die dem System schaden (z.B. über Facebook Würmer).	*Malware* – Attacking and harming computer systems via viruses or Trojans.
Q.1.6 *Cyber Bullying* – Cyber Mobbing unter Schülern.	*Cyber Bullying* – Cyber Mobbing among pupils.
Q.1.7 *Sexting* – Verbreitung von (privaten) sexuellen Fotos oder Inhalten in Facebook oder in einem anderen sozialen Netzwerk.	*Sexting* – Dissemination of (private) sexual photos or contents on Facebook or any other social network.
Q.1.8 *Happy Slapping* – Verbreitung gewaltverherrlichender (Privat-)Videos in Facebook oder in einem anderen sozialen Netzwerk.	*Happy Slapping* – Glorification of violence via (private) videos on Facebook or any other social network.
Q.1.9 *Cyber Stalking* – Verfolgung oder Belästigung einer Person in Facebook oder in einem anderen sozialen Netzwerk.	*Cyber Stalking* – Tracking or harassing of a person on Facebook or any other social network.
Q.1.10 *Cyber Mobbing* – Psychologisches Terrorisieren einer Person in Facebook oder in einem anderen sozialen Netzwerk, wobei bewusst Druck ausgeübt wird.	Cyber Mobbing – Psychologically terrorizing of a person on Facebook or any other social network by consciously exerting pressure.
b) Ja, jemand aus meinem Umfeld.	b) Yes, someone from my environment.
c) Nein, kenne niemanden, der betroffen war.	c) No, I don't know anybody who was offended.

Table A-5: Questions related to the socio-economic and socio-demographic variables $x_{2n}, \ldots, x_{n7}$ (1/2).

Questions in German	Questions in English
Q.2 **Häufigkeit**	**Frequency**
Zu Beginn denken Sie bitte ganz allgemein an soziale Netzwerke, wie beispielsweise Facebook. Als welchen der folgenden Nutzertypen würden Sie sich selbst einstufen?	At the beginning, please think of social networks such as Facebook. Which of the following user types would you consider yourself?
a) *Intensiv-Nutzer/-in* – Ich habe einen Account in zumindest einem sozialen Netzwerk und nutze diesen täglich oder fast täglich.	*Intensive user* – I have an account on at least one social network and use it at least (or almost) daily.
b) *Gelegenheits-Nutzer/-in* – Ich habe einen Account in zumindest einem sozialen Netzwerk und nutze diesen gelegentlich.	*Casual user* – I have an account on at least one social network and use it occasionally.
c) *Ehemaliger Nutzer* – Ich habe bzw. hatte einen Account in zumindest einem sozialen Netzwerk, nutze im Moment aber keinen oder ich habe alle gelöscht.	*Former user* – I have or had an account on at least one social network, but I use none or all deleted it / them.
d) *Gar-nicht-Nutzer/-in* – Ich hatte nie einen Account in einem sozialen Netzwerk.	*No user* – I never had an account on a social network.
Q.3 **Geschlecht (male / female)**	**Sex (male / female)**
Q.4 **Alter (in Jahren)**	**Age (in years)**
a) 14-19 Jahre	14-19 years old
b) 20-29 Jahre	20-29 years old
c) 30-39 Jahre	30-39 years old
d) 40-49 Jahre	40-49 years old
Q.5 Wie viele **Einwohner/innen** (EW) hat die Gemeinde oder Stadt, in der Sie wohnen?	How many **inhabitants** has the municipality or city where you live?
a) unter 10.000 EW	less than 10,000 inhabitants
b) 10.000 bis unter 50.000 EW	10,000 to less than 50,000 inhabitants
c) 50.000 bis unter 100.000 EW	50,000 to less than 100,000 inhabitants
d) 100.000 bis unter 250.000 EW	100,000 to less than 250,000 inhabitants
e) mehr als 250.000 EW	more than 250,000 inhabitants
f) weiß nicht / keine Angabe	don't know / no answer

Table A-6: Questions related to the socio-economic and socio-demographic variables $x_{2n}, \ldots, x_{n7}$ (2/2).

Questions in German	Questions in English
Q.6 Sind Sie zurzeit **berufstätig**? Was trifft auf Sie zu?	Are you currently **employed**? What applies to you?
a) Vollzeit berufstätig	Full-time working
b) Teilzeitbeschäftigt	Part-time employed
c) Arbeitslos	Unemployed
d) In Karenz	In parental leave
e) In Pension bzw. Rente	Pension, retired
f) Hausfrau bzw. –mann	House-wife or -man
g) In Berufsausbildung, Lehre, Zivil- oder Präsenzdienst	In vocational training, apprenticeship, civil or military service
h) SchülerIn, StudentIn	Pupil, student
i) Sonstiges, und zwar: _____	Other, namely: _____
j) weiß nicht / keine Angabe	don't know / no answer
Q.7 **Humankapital**	**Human Capital**
Was ist Ihre höchste abgeschlossene Schulbildung?	What is your highest completed school education?
a) Keine Pflichtschule	No compulsory school
b) Pflichtschule	Compulsory school completed
c) Lehrabschluss (Berufsschule)	Apprenticeship
d) Berufsbildende mittlere Schule	Vocational mid-level school
e) Allgemeinbildende höhere Schule ohne Matura	Grammar school without High school Diploma
f) Allgemeinbildende höhere Schule mit Matura	Grammar school with High school Diploma
g) Berufsbildende höhere Schule (HTL, HAK)	Vocational secondary school with High school Diploma (HTL, HAK)
h) Abiturientenlehrgang, Kolleg, Pädagogische Akademie	Course for high school-graduate, College, University of Teacher Education
i) Fachhochschule, Universität, Hochschule	University of Applied Sciences, Universities
j) Sonstiges, und zwar: _____	Other, namely: _____
k) weiß nicht / keine Angabe	don't know / no answer

Table A-7: Questions related to the security/incertitude variables, $x_{nS,1}, \ldots, x_{nS,14}$.

Questions in German		Questions in English
Q.S **Sicherheitsmaßnahmen**		**Security Measures**
Welche Vorkehrungsmaßnahmen setzen Sie persönlich (bzw. haben Sie persönlich gesetzt), um sich vor kriminellen Aktivitäten auf Sozialen Netzwerken zu schützen?		What security measures do you personally (resp. have you personally set) to protect yourself against criminal activities on social networks?
QS.1	Datenschutzeinstellung bei der Registrierung anpassen.	Adapt protection settings at the first registration.
QS.2	Regelmäßig Passwörter ändern.	Regularly change password.
QS.3	Auf unterschiedlichen Plattformen unterschiedliche Passwörter verwenden.	Use different passwords on various platforms.
QS.4	Sicherheitssoftware installieren.	Install security software.
QS.5	Soziale Netzwerke nicht über ungesichertes WLAN nutzen.	Do not access social networks via unsecured WLAN connections.
QS.6	Nur mit Personen in Kontakt treten, die man auch real kennt.	Only communicate with persons known in real life.
QS.7	Keine persönlichen Daten preisgeben.	Never provide personal information.
QS.8	AGBs vor Registrierung genau lesen.	Read terms and conditions carefully at every registration.
QS.9	Die automatische Passwortspeicherung deaktivieren	Deactivate automatic save password facilities.
QS.10	Cookies löschen.	Delete cookies.
QS.11	Mikrofon und Kamera des PCs überkleben.	Hide / tape microphone and camera.
QS.12	Hausverstand / Menschenverstand nutzen.	Use common sense.
QS.13	Keine Verwendung sozialer Netzwerke.	No use of social networks.
QS.14	Sonstiges: _____	Others: _____
QS.15	Gar keine.	None / I do not care about any security issues.

References

Anderson R, Barton C, Böhme R, Clayton R, Van Eeten MJ, Levi M, Moore T, Savage S (2013) Measuring the cost of cybercrime. In: The economics of information security and privacy, Böhme R (ed), Springer, Berlin / Heidelberg, pp. 265–300. DOI: 10.1007/978-3-642-39498-0.

Bailey M, Dittrich D, Kenneally E, Maughan D (2012) The Menlo Report. IEEE Security Privacy 10(2):71–75. DOI: 10.1109/MSP.2012.52.

Becker GS (1968) Crime and Punishment: An Economic Approach. Journal of Political Economy 76:1–54.

BM.I (2015) Sicherheitsbericht 2014. URL: https://www.bmi.gv.at/508/start.aspx. Accessed: 2015-10-20, publisher: Bundesministerium für Inneres, Österreich (BM.I; Federal Ministry of the Interior, Austria).

Bullée JWH, Montoya L, Pieters W, Junger M, Hartel PH (2015) The persuasion and security awareness experiment: Reducing the success of social engineering attacks. Journal of Experimental Criminology 11(1):97–115, Springer Netherlands. DOI: 10.1007/s11292-014-9222-7.

Bundeskriminalamt (2015) Polizeiliche Kriminalstatistik 2014: Sicherheit 2014. URL: https://bundeskriminalamt.at/501/start.aspx [accessed 2015-10-20]. Publisher: Bundeskriminalamt, Österreich (•BK; Federal Criminal Police Office, Austria).

Cameron AC, Trivedi PK (2005) Microeconometrics: Methods and Applications. Cambridge University Press, New York. ISBN: 978-0-521848-05-3.

Computer Professionals for Social Responsibility (CPSR) (2015) The Ten Commandments of Computer Ethics. URL: http://cpsr.org/issues/ethics/cei/ [accessed 2015-11-24].

Cook PJ, Machin S, Marie O, Mastrobuoni G (2014) Lessons from the economics of crime. CentrePiece – The Magazine for Economic Performance 18(3), Centre for Economic Performance, The London School of Economics and Political Science (LSE). URL: http://cep.lse.ac.uk/centrepiece/browse.asp?vol=18&issue=3.

Davidson R, MacKinnon JG (1993) Estimation and Inference in Econometrics. Oxford University Press, New York. ISBN: 978-0-195060-11-9.

Dimkov T (2012) Alignment of Organizational Security Policies – Theory and Practice. PhD thesis, University of Twente, Faculty of Electrical Engineering, Mathematics & Computer Science, Enschede, the Netherlands. IPA Dissertation Series no. 2012-04. ISBN: 978-9-036533-31-7 [print], URL: http://doc.utwente.nl/79740/.

Freeman R (1999) The Economics of Crime, vol. 3c, 1st edn., North Holland Publishers, Amsterdam, Netherlands, chap. 52. ISBN: 978-0-444501-89-9 [print].

Gower JC (1971) A General Coefficient of Similarity and Some of Its Properties. Biometrics 27(4):857–871, Wiley, International Biometric Society. DOI: 10.2307/2528823.

Greene WH (1997) Econometric Analysis, 3rd edn. Prentice Hall, New Jersey. ISBN: 978-0-137246-59-5.

Halaweh M, Fidler C (2008) Security perception in e-commerce: Conflict between customer and organizational perspectives. In: International Multiconference on Computer Science and Information Technology (IMCSIT), Institute of Electrical and Electronics Engineers (IEEE), pp. 443–449. ISBN: 978-8-360810-14-9 [print], DOI: 10.1109/IMCSIT.2008.4747280.

Hartel P, Junger M, Wieringa R (2011) Cyber-crime Science = Crime Science + Information Security. CTIT Technical Report Series, Centre for Telematics and Information Technology (CTIT), University of Twente, Enschede. URL: http://doc.utwente.nl/73350/.

Hinduja S, Patchin JW (2008) Cyberbullying: An Exploratory Analysis of Factors Related to Offending and Victimization. Deviant Behavior 29(2):129–156. DOI: 10.1080/01639620701457816.

Kaufman L, Rousseeuw PJ (1990) Finding Groups in Data: An Introduction to Cluster Analysis. Wiley Series in Probability and Statistics, Wiley. ISBN: 978-0-471878-76-6 [print], DOI: 10.1002/9780470316801.

Kirchner S, Angleitner B (2016) Cyber Crime. Die Social Media-Nutzer in Österreich und ihre Erfahrungen mit kriminalpolizeilich relevanten Aktivitäten. Tech. Rep., in: Wissenschaf(f)t Sicherheit. Studienband 3, Bundesministerium für Verkehr, Innovation und Technologie (bmvit), Stabsstelle für Technologietransfer und Sicherheitsforschung, Wien, pp. 45–57. URL: http://irihs.ihs.ac.at/4125/.

Kirchner S, Angleitner B, Gstrein M (2015) Cyber Crime – die Social Media-Nutzer in Österreich und ihre Erfahrungen mit kriminalpolizeilich relevanten Aktivitäten. Research Report, Institute for Advanced Studies, Vienna and MAKAM Research GmbH. Financed by the Security Research Program KIRAS of the Austrian Ministry for Transport, Innovation and Technology (unpublished).

Kochheim D (2016) Modulares Cybercrime. URL: http://www.kochheim.de/cf/ [accessed 2016-09-02].

Kshetri N (2010) The Global Cybercrime Industry: Economic, Institutional and Strategic Perspectives. Springer, Berlin / Heidelberg. ISBN: 978-3-642115-22-6.

Lee M, Crofts T, Salter M, Milivojevic S, McGovern A (2013) 'Let's Get Sexting': Risk, Power, Sex and Criminalisation in the Moral Domain. International Journal for Crime, Justice and Social Democracy 2(1):35–49. DOI: 10.5204/ijcjsd.v2i1.89.

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2015) cluster: Cluster Analysis Basics and Extensions. URL: https://cran.r-project.org/web/packages/cluster/. R package version 2.0.3.

Newman GR (2009) Cybercrime. In: Handbook on Crime and Deviance, Krohn MD, Lizotte AJ, Hall GP (eds), 1st edn., Handbooks of Sociology and Social Research, Springer, New York, pp. 551–584. DOI: 10.1007/978-1-4419-0245-0.

Saferinternet.at (2016) Das Internet sicher nutzen! Österreichisches Institut für angewandte Telekommunikation (ÖIAT). URL: https://www.saferinternet.at [accessed 2016-01-04].

Schneider C, Katzer K, Leest U (2013) Cyberlife – Spannungsfeld zwischen Faszination und Gefahr: Cybermobbing bei Schülerinnen und Schülern. Eine empirische Bestandsaufnahme bei Eltern, Lehrkräften und Schülern/innen in Deutschland. Tech. Rep., Bündnis gegen Cybermobbing e. V., Karlsruhe, Germany. URL: https://www.buendnis-gegen-cybermobbing.de/fileadmin/pdf/studien/cybermobbingstudie_2013.pdf.

Statistik Austria (2017) IKT Einsatz in Haushalten. Einsatz von Informations- und Kommunikationstechnologien in Haushalten 2015. URL: http://www.statistik.at/ [accessed 2016-07-25].

Talib S, Clarke NL, Furnell SM (2010) An Analysis of Information Security Awareness within Home and Work Environments. In: 2010 International Conference on Availability, Reliability and Security (ARES), Institute of Electrical and Electronics Engineers (IEEE), pp. 196–203. DOI: 10.1109/ARES.2010.27.

Tsohou A, Kokolakis S, Karyda M, Kiountouzis E (2008) Investigating Information Security Awareness: Research and Practice Gaps. Information Security Journal: A Global Perspective 17(5-6):207–227. DOI: 10.1080/19393550802492487.